DOCUMENT RESUME

ED 319 815 TM 015 138

AUTHOR Collins, Angelo
TITLE Novices, Experts, Veterans, and Masters: The Role of Content and Pedagogical Knowledge in Evaluating Teaching.
SPONS AGENCY Carnegie Corp. of New York, N.Y.
PUB DATE Apr 90
NOTE 19p.; Paper based on a paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Biology; Comparative Analysis; *Evaluators; High Schools; Interdisciplinary Approach; Interrater Reliability; *Knowledge Level; Peer Evaluation; Science Teachers; *Secondary School Teachers; *Teacher Evaluation; Teaching Experience
IDENTIFIERS Experts; *Rater Reliability; Stanford University CA; *Teacher Assessment Project

ABSTRACT

Since 1986, the Teacher Assessment Project (TAP) at Stanford University (California) has been exploring performance-based modes of assessment that capture the complexity of the practice of teaching. After a brief description of the rating procedures, the raters, and the situated-performances designed by the TAP for assessment, this paper describes the considerations that different classes of raters (novices, experts, veterans, and masters) use in rating teacher performance. BioTAP, the biology component of the TAP, has developed two forms of performance assessments for high school biology teachers: (1) portfolios, and (2) simulation exercises. During the 1988-89 school year, 16 high school biology teachers completed portfolios and simulation exercises; subjects were selected to represent a range of teaching experience. Sixteen teachers served as judges and rated the performance assessment activities. Among the other raters were a research biologist with little pedagogical knowledge and novices with no experience with TAP and/or teaching. Results indicate that raters interpreted teaching tasks in terms of their own backgrounds, with each class of raters bringing a specialized knowledge to the task. It seems that the most valid rating system would involve a multidisciplinary team of raters. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319815

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Angelo Collins

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Novices, Experts, Veterans, and Masters: The Role of Content and Pedagogical Knowledge in Evaluating Teaching

Angelo Collins
Stanford University

Paper Based on a Presentation at the
Annual Meeting of the
American Educational Research Association
Boston, Massachusetts
April 19, 1990

Running Head: Novices, Experts, etc.

The research reported here was supported by a grant from the Carnegie Corporation of New York. The opinions expressed are those of the author and in no way reflect the views of that organization or the National Board for Professional Teaching Standards.

For further information on the Teacher Assessment Project write: Teacher Assessment Project, CERAS 507, Stanford University, Stanford, CA 94305-3084

TM015138

BEST COPY AVAILABLE

Novices, Experts, Veterans, and Masters: The Role of Content and Pedagogical Knowledge in Evaluating Teaching

If teaching is a profession, or an art, or a craft, then the proper assessment of teaching should not be analogous to meeting the specifications of piece work done by a technician. Rather the assessment of teaching should include recognition of the technique, the performance and the product analogous to, but much more complex than, professional ice skating, for example. Since 1986, the Teacher Assessment Project (TAP) has been exploring performance-based modes of assessment that capture the complexity of the practice of teaching. One of the assumptions of TAP has been that teaching occurs in a context -- something is taught to someone at some point in time. After a brief description of the situated-performances designed by the Project for assessment, the rating procedure and the raters, this paper will describe the considerations that different classes of raters -- novices, experts, veterans and masters -- used what rating teacher performance.

The Assessments

BioTAP, the biology component of TAP has developed two forms of performance-based assessment for high school biology teachers: portfolios and simulation exercises. A portfolio is a collection of evidence that allows teachers to demonstrate their solutions to teaching problems. Portfolios were expected to be especially useful in capturing the context of teaching and the growth and development of both students and teachers that is an essential part of teaching. In the BioTAP research, the biology teacher's portfolio was divided into five sections, called entries. Teachers completed an entry that contained background information on their education, experience, and current teaching assignment. This entry was meant to inform a portfolio rater of the context of the evidence, and, as such, was not rated. A second entry contained evidence about how the teacher planned a unit, including the plan itself. A third entry contained evidence about how the teacher taught a lesson doing either laboratory work or using a source other than the text book.

This instruction entry had a videotape of the teachers in their instructional setting. The fourth entry focused on how teachers assessed students and included samples of student work as evidence. The final entry allowed the teachers to focus on their professional development and professional service (Bird & King, 1989).

The other form of performance-based assessments used by BioTAP was simulation exercises. Simulation exercises provide an opportunity for teachers to perform critical tasks of teaching in a standardized situation.. The simulation exercises at the assessment center for high school biology teachers were of two types -- some were situated in the portfolio while others were independent. One of the exercises situated in the portfolio was a unit plan review that emphasized how student diversity influenced planning. Another exercise probed teachers about their rationale for the assessment of students. A third used the laboratory teaching portfolio entry to investigate teachers knowledge and skill in classroom management and content knowledge. The last portfolio-based assessment center exercise was an extension of the portfolio entry on using materials for instruction other than the textbook. The independent simulation exercises included one on using a computer as an instructional tool and another on adapting a textbook to local circumstances. The third independent exercise provided teachers the opportunity to critique a videotape of other teachers while reflecting on their own teaching. The last independent exercise was group-administered and focused on the problems of teaching evolution (King, 1989).

The Teachers

Sixteen high school biology teachers completed portfolios during the 1988-89 school year and completed the required simulation exercises in an assessment center during June, 1989. These teachers were not random but were selected for years of teaching experience -- from an intern to a veteran with 29 years experience. Males and females were equally represented in the sample. The teachers were members of different ethnic groups and taught in schools with a variety of ethnic compositions. In all, each teacher completed 11 separate performance-based assessment activities: four portfolio entries, three portfolio-based simulation exercises and four independent simulation exercises.

The Rating Procedure

If portfolio entries and simulation exercises are meant to capture the performance of the complex practice of teaching, it is important to consider what are appropriate methods to rate such performances and who is qualified to do such rating. The evaluation of ice skating performance is accomplished while the performance is taking place and the judges are seldom practicing ice skating professionals. Rather, the judges are former skaters, coaches, and others who have developed a keen eye for quality performance. In an attempt to begin to look at answers to the questions of how to rate performance-based teacher assessments and who is qualified to do such rating, BioTAP developed a rating procedure that placed several constraints on raters to score the performance in a holistic manner.

First, the rating form required raters to judge the performance of the teacher in six categories. The first five categories were modifications of the core propositions of what teachers should know and be able to do, developed by the National Board for Professional Teaching Standards (1989). The rating categories were:

1. The candidate attended to students and their learning;
2. The candidate knew the subject matter and how to teach it;
3. The candidate attended to class management and monitoring;
4. The candidate thought about and learned from his/her activity;
5. The candidate participated in a learning community; and
6. Overall rating.

A second constraint placed on the raters was that they had to use a a scale of zero to five to rate the goodness of the teaching performance. A score of zero indicated that the rater was unable to form an opinion; a score of one indicated the judgement was that the performance was unacceptable; a score of two indicated a weak performance; a score of three indicated an adequate performance; a score of four indicated a proficient performance; and a score of five indicated a judgement of a superb performance. On the assumption that teachers attempting these performance-based assessments would generally be adequate, the score for each performance was entered as three.

Another constraint that pushed raters to making a holistic rating was the design of the

rating form. Although the procedure required raters to state a rationale each time the default value of three was changed, the space for comments was less than a half page.

The final constraint placed on the raters was time. Sixteen raters were required to complete almost all of the rating of eleven assessment activities (both portfolio entries and simulation exercises) for sixteen teachers in less than two days. Further, it was required that each performance be rated twice by different raters so that differences between raters might be considered. Some of the portfolios were rated prior to the two days assigned to rating and some of the simulation exercises were rated as the exercise was being administered. Another time constraint was in the training received on rating procedures. Although each rater was given a rating manual that described the procedure, the actual training in its use was less than four hours.

Time, one of the factors that served as a constraint to force the raters to look at the performance of the teacher as whole, contributed to one of the weaknesses of the rating process. The opportunity to become proficient either on rating a single teacher on all eleven activities or on rating the same activity for all sixteen teachers was sacrificed to complete the rating task in a limited time. There were two other weaknesses in the rating procedure. One was that all raters were, of necessity, beginners at using the rating system. The other was in the attempt to double-rate all performances. In the rating of simulation exercises, some raters were able to rely on original sources, as they administered the exercise, while other raters were required to rely on second-hand sources, either reading the notes taken by the exercise administrator or by listening to an audiotape of the performance. With the constraints, and despite the weaknesses, the rating task was completed in a timely manner and judgments were made about the quality of the performance.

The Raters

Altogether, sixteen persons were engaged in the task of rating the performance-based assessment activities by the sixteen high school biology teachers who participated in BioTAP. Compelled by the assumption that the assessment of teachers is best accomplished by teachers, and that the subject matter context of the teaching is important, it

was decided that most of the raters should be experienced high school biology teachers. This group of nine teachers constituted the masters referred to in the title, persons assumed to rank high in both content knowledge and pedagogical expertise. Among the other raters, there was one expert, a research biologist, assumed to be have deep and extensive content knowledge in biology, but relatively little pedagogical knowledge. This rater did claim knowledge of pedagogy because of experiences teaching graduate students in biology. Two of the raters were classified as novices. Neither had experience in teaching biology nor extensive course work or research experience in biology. Therefore it was assumed that they were low in both content knowledge and in pedagogy. However, neither were true novices as they had both been active in the Teacher Assessment Project and have conducted research on good teaching and its assessment. One of the raters was classified as a veteran, an experienced high school English teacher, assumed to have little content knowledge in biology but much pedagogical knowledge. Despite the title of this paper, there was a fifth group of raters, possibly called coaches. These raters were three university faculty with degrees and experience in science education. All have experience in science research, in teaching, in teaching teachers, and in educational research. However, one has extensive experience in physics while the other two have done their science work in biology.

As the samples were small, a numerical analysis of scores by raters in different classes is not fruitful. Rather, to identify patterns within and across different classes of raters requires an examination of the comments that the raters made to accompany their scores. The description of the raters comments will begin with novices, followed by the veteran and the expert respectively. Then comments by the science educators will presented. As the largest class, the last of rater comments to be presented will be the masters. The comments will be grouped according to the categories for rating the teachers' performances.

Rating Results

The Novices

The two novices rated a total of 41 performances. More than half of this rating was

done on the personal and professional development portfolio entry. The remainder of the ratings were done randomly and included the alternate materials portfolio entry, the laboratory lesson portfolio entry, the unit plan exercise, the videotape reflection exercises and the textbook exercise. Each of the novices rated at least one performance by each teacher. The most noticeable characteristic of the comments made by the novices is their descriptive nature. For example, one comment reads, "[Teacher] describes a range of contacts and activities, but none in much detail. His coordination of biology together with need to cooperate with coordinators of other [science] subjects suggests extensive contact within the department." Another comment reads, "[Teacher] is department chair, and has a range of professional activities in department, and school, and outside it."

The novices did not hesitate to address issues related to teachers being attentive to students and their learning. Comments were made when a teacher was attentive to, or ignored the needs of minority students in the class. In other instances related to students and their learning, the novices commented that the teacher

> "had never referred to a single student by name,"
>
> "did not follow-up on student questions,"
>
> "did not seem concerned with student interests only [his] own goal," and
>
> "suggested eliminating classroom diversity by creating an artificial homogeneity through tracking."

Although there is no pattern in these comments, they do demonstrate that the novices were able to identify teachers' concern (or lack thereof) with students and their learning.

With regard to understanding the subject matter and how to teach it, with one comment as an exception, novices avoided reference to subject matter competency. The one exception is a comment that the "teacher seems to have average subject matter knowledge as he accepted the text uncritically." However, the novices did not hesitate to comment on pedagogy and did so in two types of comments: either about general pedagogical technique or by placing themselves in the position of students. In the first instance, one typical comment was " a technically competent lab -- introduction, the teacher modeled what to do, the students did it, and there was a debrief." In the latter instances, a sample comment was

"The bad [bacteria] culture seemed to be a nuisance to the teacher, rather than an opportunity. Why did it go bad?" With regard to content specific pedagogy, the novices' comments included their own uncertainty. For example, "It appeared to me that this lab required lots of preparation, gathering materials, but seemed disorganized and designed to keep the students busy."

In the category of classroom management, comments again were not specific to a biology classroom, for example, "Students knew what to do and teacher monitored them." In the category of evaluation and reflection the comments of the novices were more perceptive than in the other rating categories. For example, a comment that accompanied an evaluation statement by a novice about the technical competence of a laboratory activity was that "[the teacher] wrote reflective comments on operation of the laboratory, but did not reflect on the the value of the experience." Similarly, novices expected that reflective statements would have a depth.

As expected from the assessment activities rated by the novices, the majority of the comments made by the novices were about teachers' participation in a learning community. The amount and type of reflection by the teacher was a feature in all of the comments, which were remarkable similar. They all were directed at the amount of professional activity and the teacher's awareness of the purpose of the activity. Typical comments include"lots of contacts, but no reason why," "lots of activity with no plan, no reflection, no rationale," "one sustained professional contact with great enthusiasm and no critical reason," and " systematic, goal-oriented."

From these comments it is reasonable to assume that novices with respect to biology teaching can make judgements about assessment activities intended to assess biology teachers, but only in certain categories or in assessment activities designed to assess certain types of knowledge and skills. The comments made about students and their learning seem sufficiently on target and similar to comments made by other raters to be appropriate. Novices avoided comments about subject matter. Therefore, information about a teacher's subject matter knowledge would not be adequately assessed by novices. If a biology

teacher were teaching incorrect factual material, it would be undetected by novices. It appears that these novices were able to assess general pedagogy, but were unable to discern evidence of content specific pedagogy. For example, in the technically adequate laboratory experience, what safety measures were taken. It is consoling that in judging professional development activities, the novices looked for rationale and pattern, rather that quantity. However, as non-biology teachers they did not attempt to discern if different professional development activities were qualitatively and quantitatively different from each other.

The Veteran

The veteran teacher rated 19 performances by 14 of the 16 teachers. All but two of the ratings were done on either the unit planning portfolio entry or the unit planning review exercise. The remaining two ratings were done on the textbook exercise. The veteran teacher concentrated the rating comments on four rating categories, omitting any comments on managing and monitoring the classroom. The first category addressed was students and their learning. In every rating instance a reference was made to whether or not the teacher provided evidence of concern for students. Typical comments were "Sensitive to limits of students," and "Sees students as stereotypes." In the category of knowledge of subject matter and how to teach it, the veteran teacher referred either to general pedagogical knowledge, especially variety in instruction or pedagogical knowledge related to teaching reading, writing or literature. Comments typical of the former are, "lack of creativity, limited to book and worksheets," and "no evidence of students doing anything but listening to lectures, looking at overheads and reading." Comments typical of the latter are "Creative use of role-play," and " unaware of reading problems." The veteran made two references to subject matter knowledge: "Subject matter knowledge seems confused," and " Since [teacher] has no experience in this field, relies on textbook."

As with the novices, the veteran commented frequently on the teachers being reflective, for example, when giving reasons for decisions about instructional strategies on unit plans. Reflection was also associated with identifying student needs. In both instances where the evidence reviewed was a unit plan on the topic of biotechnology, the veteran mentioned that it was a shame the teacher taught in Silicon Valley and had not in any way relied on the

resources available in such a technology-rich environment.

In many ways the veteran was like the novices. Most of the rating comments were general, and the comments focused on what the veteran knew well -- students, teaching strategies, reflection and rationale. Therefore, the same difficulties in providing a valid rating would be likely if the only criteria for rating performance-based assessments was experience in teaching. -- subject matter knowledge and content-specific pedagogy would be lost.

The Biologist / The Expert

The biologist rated fourteen performances equally divided between the student assessment portfolio entry and exercise, the laboratory exercise and the computer exercise. These fourteen ratings were of performances by nine teachers. The most noticeable characteristics of the comments written by the biologist is their length -- they are more than three times longer than any other set of comments by any other rater. The comments contain both rich descriptions of the performance and qualifications about the performance. In each rating instance the biologist made a comment about how concerned the teacher was about students and their learning. Most of these comments were positive and many were typical of other raters. Two examples include "adjusts lesson to language problems" and "assigns tasks that students can achieve to build student self esteem." Two of the negative comments were: "students seem to have much difficulty with learning and [the teacher] is always blaming it on the students," and "[The teacher] hasn't considered their [the students] social development."

For the biologist, the comments on the subject matter knowledge of the teachers was always tied to the teaching. Several examples follow:

> -- "Is extremely knowledgeable about genetics, even to knowing the common misconception and so has easily created multiple paths to the knowledge;"
> -- "has not taught critical thinking skills along with the content."
> -- "Knows [difficult] concepts but doesn't seem to be reaching students."
> -- "Modest evidence that [teacher] knew subject matter. More evidence that [teacher] knew teaching methods and skills."

The biologist also commented that "While the exercise is designed to give ample

opportunity to show how much [the teacher] knows about the subject (genetics), the response didn't take the opportunity."

There were relatively few comments by the biologist on classroom management and they were all vague, such as "able to monitor well." There were few comments on reflection but all of them noted that the teacher had either given a reason or had begun an action based on the reflection. Examples include "is critical of her own examination and learned from the experience" and "is aware of the problem (of attendance) and is looking at its causes or a cure." In three instances the biologist made comments about a teacher being a member of a learning community -- about working with another teacher, about working with industry, and about not seeking information from a colleague.

The biologist's responses with regard to subject matter were unanticipated -- not made in isolation of teaching. It had been expected that she would look sharply at how accurate and how current the content knowledge was. The comments about the teaching were consistent with those made by the masters. It was not surprising that the biologist did not comment on classroom management. Comments about the other performance categories were perceptive. It may be that the exercises that the biologist rated were not the best to allow teachers to express their subject matter knowledge, yet, as the biologist said, "there was ample opportunity [to talk about content] and the teachers did not respond."

The Science Educators

The three university science education faculty members rated a total of 68 performances on all assessment activities except the alternative materials activities, and in the process judged materials from all the teachers. The comments made by the science educators were broad, judgmental claims, uniformly terse. Frequent comments in all categories were, "nothing exceptional," "adequate," "normal," "expected," and " no glaring errors but nothing outstanding."

Although comments were made by the science educators about the teachers' concern for students and their learning, they were less common than in any other group. When comments were made, they were always to identify a flaw. For example, "Concern for students does not translate into action, " or "students seem absent from this unit plan." One

striking comment was "Does not evaluate student knowledge - rather completion, submission, and attendance are major factors."

The science educators were explicit in addressing the teachers knowledge of subject matter. Comments identify adequate command of subject matter included, "command of subject matter sound," and "outstanding grasp of subject." However, they were just as quick to counter with examples where subject matter knowledge was weak. For example, "very little subject matter knowledge evident," and "subject matter knowledge limited and superficial." In addition, these raters pointed out specific and general errors in content: "[teacher] misused 'population' and 'community' throughout the discussion," "does not know the difference between 'gene' and 'allele,'" "subject matter knowledge out of date," and "improper use of terms on the worksheet." The science educators were also aware of rating how the teacher taught the subject matter, that is content specific pedagogy. For example:

-- "good use of analogies that would appeal to students,"
-- "impressive knowledge of subject and how to teach it,"
-- "plans are realistic and pragmatic for amount of content and labs,"
-- "uses analogies but they are not linked to students or subject matter,"
-- "good balance of teaching strategies appropriate to subject."
-- "understand nature of science and plans variety of activities accordingly."

Reflection was commented on frequently by the science educators. Examples include "much self- monitoring," "needs more rationale," "very analytical," "rich and thoughtful," "no reason, no reflection," and "very wise." When the science educators commented on the teachers as members of learning communities, the comments echoed those of the novices, looking for reasons and patterns. For example: "lots of relationships -- no pattern, no reason, no goal," "No evidence that collaboration leads to substantive change in science knowledge." and "[Teacher] could ask advice from more experienced teachers."

It may be that science educators did not comment on concern for students and their learning because they view it as a sine qua non of teaching and therefore only comment on its absence. It may be that these raters criticized subject matter knowledge and content specific pedagogy because of their experiences in observing and coaching beginning

teachers. As the novices, they looked for patterns and reasons in reflection of the teachers and in their professional activities.

The Masters

Nine experienced high school biology teachers constituted the group of masters. Three had received either state or national awards for excellence in teaching biology. One of the masters, although an experienced teacher, does not have a license, has taught in a private school, and has never taken a formal teacher education course. Only one had less than ten years experience teaching biology. These nine raters scored 195 performances covering all the assessment activities and all the teachers. Several of these raters developed their own vocabulary and/or criteria for rating. These included being "well-educated," biologically literate," and "creative." There were also several phrases that the masters used with regularity that were adapted from the raters manual. These included "superficial and general," "adequate," "nothing outstanding" and "nothing new and different from common practice." Three raters consistently commented on grammar and spelling errors in portfolio entries. The harshest comment was terse, "Entry was vacuous."

In every set of comments by every master about every teacher there was a phrase about how well the teacher attended to students and their learning. These comments were of three types: 1) recognizing the attention to students; 2) noting the absence of attention to student needs and differences; and 3) noting a discrepancy between what the teacher claimed and what the rater noted in the performance. Typical of the comments that indicate awareness of attending to students' learning were:

-- "aware of levels of ability of students,"
-- "has written own text because of student language deficiencies,"
-- "well attuned to the feelings and insecurities of adolescents,"
-- "many techniques to make knowledge available to students with different needs,"
-- "knows students behavioral and learning needs and has developed many quick, effective techniques to engage them,"
-- "great flexibility of what is allowed and expected."

Some of the comments that indicated that the teacher was not aware of students and their learning were:

-- "unaware that students have different needs,"
-- "I cannot get a picture of hat the kids are like and are doing in this person's class,"
-- "no evidence of tailoring to meet student needs,"
-- "already knows what is best so is not receptive to students ideas,"
-- "only attentive to those who show initiative."

Comments about teachers attention to students and their learning that were unique to the masters were the comments about the discrepancy between the teacher's claim and the observed performance. For example, masters said:

-- "tries to make knowledge accessible to all, but because of emphasis on addressing needs of minority students, some students not attended to,"
-- "great variety for strategies, but nothing spoke to relationship of strategies to differences between students and their learning,"
-- "knew about students ethnic background and felt it helped him teach, but not able to say how."
-- "most emphasis was on student needs but goal is to keep [students] busy, not engage their minds."

In the rating category about knowing the subject matter and how to teach it, comments by the masters focused on the subject matter, teaching it, and the intersection of the two in content specific pedagogy. Comments on subject matter knowledge were, in most instances, general. For example, while praising teachers the masters said "very good knowledge of subject," "was up on the most recent research," and "saw science as a body of knowledge with themes." Comments that indicated a deficiency in science knowledge included "seems to miss the point himself of what he is teaching," and "sees science as facts and figures, not as processes." With one exception, masters made no references to errors in content knowledge, and that one reference was not specific, "errors in use of terms." Comments were also made about the teachers integrating subject matter knowledge from chemistry, writing, reading, sociology and history into biology instruction. For example, used knowledge of history to support ideas." Comments that were directly related to pedagogy but not to the teaching of biology included

-- "[teacher] only values recall of information supplied by the teacher,"
-- "consciously teaches higher order thinking skills,"
-- "[teacher] says [she] uses variety but only evidence of word games and worksheets."

-- "includes application and extension questions on worksheet,"
-- "has wide variety of strategies."
-- "engaging lecturer."

Comments on the intersection of biology knowledge and teaching, content specific pedagogy, range from vague -- "knows subject matter but not how to teach it," -- to specific --"very well informed, spotted errors in what the lab claimed to do and what it could do." Others comments on content specific pedagogy included

-- "wide variety of methods but not hooked to the topic,"
-- "dry, dull, repetitious - no sense of overall scheme of biology,"
-- "really knows how to prep a lab,"
-- "feels the process of learning is equally and maybe more important than leaning terminology and content."

There were very few comments from the masters in the rating category of managing and monitoring the classroom, and these were trite. For example " well-managed," and "only uses eye contact." Similarly, with the exception of one master who commented on the reflections made by every teacher, there were few comments about reflection. And again with one exception, the comments that were made by the masters as a whole did not indicate an expectation that teachers would express reasons for what they were doing. Comments on reflection were more general, such as: "brave and risky behavior," "constant revision of lesson plans indicates reflection," and "needs to think more as his practice does not reflect his good intentions." One eloquent comment made by a master indicates that he believes that teachers should be able to reflect, "[teacher is] only able to talk about 'what,' not 'why' and 'how.'" There were also very few comments made by the masters about teachers as members of a learning community. The comments include

-- "must have, because team teaches,"
-- "must have to be so creative,"
-- "uses local park,"
-- "uses the library,"
-- "uses the special education staff,"
-- "actively engages parents."

The masters, current high school biology teachers, placed most emphasis in their rating on the two activities that are most likely to take place in the classroom -- student learning

and teaching biology. They placed relatively little emphasis on the assessment activities that do not capture teaching as it is currently practiced -- reflection and participation in learning communities. The one master who commented on reflection for every teacher has served as a supervisor in a teacher education program where reflection is emphasized. The masters were the only group of raters to match what teachers said they did and what the evidence indicated they actually did.

Conclusions

With the exception of classroom management and monitoring, all classes of raters made comments in all the rating categories. Rather than concluding that none of the raters were concerned about classroom management or that no teacher performance in this category was noteworthy, the fact that there were so few comments might be explained by the fact that no assessment activity was specifically designed to capture classroom management. It had been assumed that the raters would make inferences about performance in this category from other evidence, particularly the videotape of the teacher in the classroom.

In general, despite a little bit of training and a common rating manual, the raters interpreted the task in terms of what they are accustomed to doing. The novices looked for what they knew about teaching. Therefore, while their comments were not inaccurate, they emphasized patterns and generalities while avoiding content specific pedagogy about biology. The veteran also emphasized what she knew, while being very careful to avoid what she knew she didn't know. Therefore, in her comments, statements about discipline specific knowledge were conspicuously absent. The science educators also did what they usually do -- look for characteristics of a performance that they can coach and that are consistent with current teacher education practice. Therefore, they saw content errors, which can be corrected; they saw content specific pedagogy, and they looked for reasons and patterns in actions and words. The categories that the masters emphasized also is consistent with current practice. They placed their emphasis on students and on teaching. The biologist is an exception to the generalization that the raters do what they are accustomed to doing -- she did not comment on subject matter knowledge. This might be explained by the fact that as an expert, she holds this knowledge tacitly. Her emphasis on

the relationship between content and knowledge might be explained by her socialization by the Project, or maybe by some predisposing concerns that made her interested in the Project in the first place.

As to the question of who should and could rate performance-based assessments of teaching, it would seem that each class of raters brought specialized knowledge from their own experiences to the task. If practical considerations, such as cost and organization, were not a factor, the most valid rating would be by teams made up of teachers in and out of their own content area, teacher educators and content specialists. However, as this recommendation is not efficient it does not seem likely it will be implemented. Therefore, it is recommended that classroom teachers be given opportunities to participate in experiences that make them more proficient raters. On such experience might be for a group of teachers to discuss the rating categories, observe each other teach, and discuss what happened and why. This activity would provide experiences in reflection and in developing a learning community.

Collins: Novices, Experts, etc.

References

Bird, T. & King, B., (1989). Report of the use of portfolios to assess biology teachers. Stanford University, Teacher Assessment Project.

King, B. (1989). Assessment center exercises for high school biology teachers. Stanford University, Teacher Assessment Project.

____(1989). Toward high and rigorous standards for the teaching profession. Detroit, MI. National Board for Professional Teaching Standards.